Am. J. Hum. Genet. 66:235–250, 2000

High Levels of Sequence Polymorphism and Linkage Disequilibrium at the Telomere of 12q: Implications for Telomere Biology and Human Evolution

Duncan M. Baird,* Joanna Coleman, Zoë H. Rosser, and Nicola J. Royle

Department of Genetics, University of Leicester, Leicester, United Kingdom

Summary

The human Xp/Yp telomere–junction region exhibits high levels of sequence polymorphism and linkage disequilibrium. To determine whether this is a general feature of human telomeres, we have undertaken sequence analysis at the 12q telomere and have extended the analysis at Xp/Yp. A total of 22 single-nucleotide polymorphisms (SNPs) and one 30-bp duplication were detected in the 1,870 bp adjacent to the 12q telomere. Twenty polymorphic positions were in almost complete linkage disequilibrium, creating three common diverged haplotypes accounting for 80% of 12q telomeres in the white population. A further 6% of 12q telomeres contained a 1,439-bp deletion in the DNA flanking the telomere. The remaining 13% of 12q telomeres did not amplify with the primers used (nulls). The distribution of telomere (TTAGGG) and variant repeats within 12q telomeres was hypervariable, but alleles with similar distribution patterns were associated with the same haplotype in the telomere-adjacent DNA. These data suggest that 12q telomeres, like Xp/Yp telomeres, exhibit low levels of homologous recombination and evolve along haploid lineages. In contrast, high levels of homologous recombination occur in the adjacent proterminal regions of human chromosomes. This suggests that there is a localized telomere-mediated suppression of recombination. In addition, the genetic characteristics of these regions may provide a source of deep lineages for the study of early human evolution, unaffected by both natural selection and recombination. To explain the presence of a few diverged haplotypes adjacent to the Xp/Yp and 12q telomeres, we propose a model that involves the hybridization of two archaic hominoid lineages ultimately giving rise to modern *Homo sapiens*.

Received August 20, 1999; accepted for publication November 4, 1999; electronically published January 11, 2000.

Address for correspondence and reprints: Dr. Nicola J. Royle, Department of Genetics, University of Leicester, University Road, Leicester, LE1 7RH, United Kingdom. E-mail: njr@le.ac.uk

* Present affiliation: Department of Pathology, University of Wales College of Medicine, Cardiff, United Kingdom.

0002-9297/2000/6601-0026$02.00

Introduction

Telomeres are tandemly repeated structures that cap the ends of linear eukaryotic chromosomes, protecting them from end-to-end fusion and degradation (Blackburn 1991). In humans, telomeres consist predominantly of the sequence TTAGGG, which is reiterated in arrays of ⩽20 kb (Hastie et al. 1990). The telomere-repeat motif is synthesized de novo at the terminus by the ribonucleoprotein telomerase; this counteracts the loss of terminal sequences resulting from incomplete lagging-strand synthesis. The length of the telomeric arrays depends on both the telomerase activity and the replicative history of the tissue in which it resides (Harley et al. 1990). Proximal to the telomere-repeat arrays are regions containing tandemly repeated DNA-sequence families (reviewed in Royle [1995]). These are shared between nonhomologous chromosomes with a polymorphic distribution between unrelated individuals (de Lange et al. 1990; Ijdo et al. 1992; Royle et al. 1992). For example, detailed analysis of the terminal region of chromosome 16p revealed extraordinary length polymorphism in the subtelomeric region. Four different-length alleles were identified, but the telomere-adjacent sequence of these alleles tended to show more similarity to subterminal sequences on other chromosome ends than they did to one another (Wilkie et al. 1991).

Human chromosome ends (proterminal regions) display elevated levels of recombination, as is evidenced by chiasmata distribution and genetic-map expansion (Hulten 1974; National Institutes of Health/CEPH 1992; Mohrenweiser et al. 1998). In addition, the terminal regions of primate chromosomes have undergone rapid and large-scale changes over short periods of evolutionary time. All of the gorilla and half of the chimpanzee chromosome ends contain a subterminal satellite sequence that is absent from both the orangutan and human chromosome ends (Royle et al. 1994). Analysis of sequences orthologous to the human Xp/Yp-telomere junction in the chimpanzee, gorilla, and orangutan genomes revealed rapid sequence evolution involving gross chromosomal rearrangements; consequently, the position of the telomere is unique to the *Homo* lineage, with

additional structures being present in the other species examined (Baird and Royle 1997). All of these observations suggest that, in the great apes, chromosome ends are dynamic and relatively transient structures, such that the majority of human telomeres have been created since the *Homo*/*Pan* divergence (Royle et al. 1994; Baird and Royle 1997).

To examine repeat turnover within specific human telomeres, we have developed a system (telomere-variant-repeat mapping by PCR [TVR-PCR]) with which to assay the distribution of variant repeats that occur within the proximal 1.5 kb of human telomere-repeat arrays (Allshire et al. 1989; Baird et al. 1995). The results of TVR-PCR analysis of two human telomeres (Xp/Yp and a subset of telomeres on chromosome 16) have revealed a hypervariable distribution of repeats (Baird et al. 1995; Coleman et al. 1999). Comparison of allelic telomere codes suggests that these repeat arrays have evolved along haploid lineages and that the mutational mechanisms underlying the observed variation are predominantly of an intra-allelic nature, involving small localized expansions/contractions in blocks of like repeat units. This may involve mechanisms such as replication slippage and unequal sister-chromatid exchange (Baird et al. 1995). Analysis of the Xp/Yp telomere–adjacent DNA revealed a high frequency of sequence polymorphism clustered within the 1 kb immediately adjacent to the telomere. In addition, strong linkage disequilibrium across the polymorphic sites results in a limited number of highly diverged haplotypes, with a striking absence of intermediate haplotypes in the population examined (Baird and Royle 1997; Baird et al. 1995).

To determine whether the unusual genetic characteristics of the human Xp/Yp telomere are unique or are a general feature of human telomeres, we have analyzed sequence polymorphism at the telomere of 12q. In addition, we have extended the analysis of the Xp/Yp terminal sequences in African DNA samples.

Material and Methods

Genomic DNAs

The samples from white individuals were from the CEPH collection of lymphoblastoid DNAs. African DNA samples (AF) were extracted both from whole-blood samples collected in the United Kingdom and from semen samples that were kindly supplied, to Professor Sir Alec J. Jeffreys, by A. D. Nkomo and S. B. Kanoyangwa (Forensic Science Laboratory, Causeway, Zimbabwe). All the studies conducted had obtained prior approval from the Leicestershire Health Authority's Committee on the Ethics of Clinical Research Investigation.

PCR

PCR was done with the use of the buffer system described elsewhere (Jeffreys et al. 1991). The primers and annealing temperatures used for PCR were as follows: TSK8B 5′-CCCTCTGAAAGTGGACCTAT-3′ at 58°C, 12qB 5′-ATTTTCATTGCTGTCTTAGCACTGCAC-3′ at 61°C, TSK7B 5′-CACTATTAGGGTTATTATGTTGACTA-3′ at 60°C, 12qA 5′-GGGACAGCATATTCTGGTTACC-3′ at 61°C, 12qArev 5′-CCCCAAAATATAATGGTAGGTAACC-3′ at 63°C, pKSRV2.D 5′-ATCCTAGCAAAGCTGAGAACTCAG-3′ at 63°C, 12q-197A 5′-GGGAGATCCACACCGTAGCA-3′ at 67°C, and 12q-197G 5′-TGGGAGATCCACACCGTAGCG-3′ at 67°C. In addition, primer pKSRV2.G 5′-CACAGTAGACAAGGGTAAGGTTTG-3′ was used for sequencing of some PCR products, and primer 12qnull3 5′-GATGTCTGAGTGGATTCAGACATG-3′ was used for analysis of 12q-deletion-allele–associated telomeres.

TVR-PCR was performed as described elsewhere (Baird et al. 1995). Individuals who were selected for TVR-PCR analysis were heterozygous (either A,B or A1,B) for haplotypes in the 12q telomere-adjacent DNA, or they were hemizygous, carrying only one amplifying allele. Allele-specific amplification for TVR-PCR analysis of the 12q telomere was conducted from the −197 polymorphism, by use of primers 12q-197A or 12q-197G. Products of TVR-PCR were resolved on a 5% polyacrylamide gel, 1 × Tris-borate EDTA, and 7.67 M urea, at 50°C for ~4 h, 30 min.

Polymorphism Assays

Polymorphic sites −1036 (*Taq*I+/−) and −1812 (*Alu*I+/−) were assayed after PCR amplification with primers pKSRV2.D and 12qArev was done under the following conditions: denaturation at 96°C for 20 s, annealing at 63°C for 30 s, and primer extension at 70°C for 80 s for 32 cycles. PCR product, in amounts of 1 μl and 5 μl, was digested with *Taq*I and *Alu*I, respectively, according to the manufacturers' recommendations, and the products were resolved on 2.5% and 4.5% Metaphor agarose gels (FMC BioProducts), respectively.

Polymorphic sites −63 (*Aci*I+/−), −473 (*Bsm*I+/−; fig. 1), and −554 (*Kpn*I+/−), and the −(659–630) insertion/deletion (in/del) were assayed after PCR amplification was done, with primers 12qA and 12qB or TSK7B, under the conditions described above but with an annealing temperature of 61°C. A total of 3 μl of the PCR reactions was digested with the appropriate restriction enzyme, and the products were resolved as follows: *Aci*I, on 3.5% Metaphor agarose gel; *Bsm*I and *Kpn*I, on 1.5% agarose gel (high-gelling-temperature agarose; FMC BioProducts) (see fig. 1).

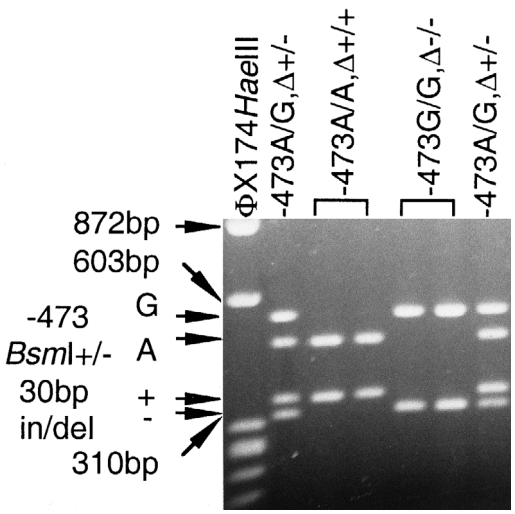

Figure 1 An example of a *Bsm*I RFLP assay. This assay identifies both the −473 SNP (upper two fragments) and the −(659–630) insertion/deletion polymorphism (lower two fragments). Note the association between the polymorphisms observed in DNAs that are either homo- or heterozygous for both.

Sequencing

Gel-purified PCR-amplified double-stranded DNA was sequenced by use of the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems). Contiguous sequences were assembled by use of the Autoassembler program, version 1.4, for Macintosh (PE Biosystems). Sequence analysis of 28 alleles in the telomere-adjacent DNA was conducted on amplicons generated with primers 12qB and pKSRV2D, to examine the extent of linkage disequilibrium. Haplotype A alleles (9) were obtained from four DNAs that were homozygous for haplotype A and from one that was hemizygous (A/null); haplotype B alleles (10), from four DNAs that were homozygous for haplotype B and from two that were hemizygous (B/null); and haplotype A1 (9) alleles, from two DNAs that were homozygous for haplotype A1 and from five that were hemizygous (A1/null). Sequence analysis of telomere repeats was performed after the excision of TVR-PCR bands from the polyacrylamide gel, as described elsewhere (Baird and Royle 1997), and was followed by reamplification with both the allele-specific primer used in the original amplification and the TAG primer (Jeffreys et al. 1991).

Results

Sequence Analysis of the 12q Telomere–Adjacent DNA

The 12q telomere–adjacent DNA was isolated by virtue of homology with the telomere-adjacent DNA of chromosome 7q (Royle et al. 1992; Coleman and Royle 1996). A small amount of sequence information (74 bp) was available from the 12q telomere–junction clone TSK7 (Royle et al. 1992). The sequence showed 91% similarity to the 7q telomere–adjacent sequence, and the sequence differences were used to design a 12q-specific PCR primer (TSK7B). Combinations of TSK7B and other primers from the 7q sequence were used to amplify 12q telomere–adjacent sequences from a half-YAC clone, yRM2196 (gift from H. Riethman), containing the terminal regions of 12q. Amplicons were sequenced, and a contig of 1.9 kb was determined from 12q. Additional 12q-specific primers (12qA and 12qArev) were designed, and the sequence was verified by amplification of genomic DNA. The chromosomal location was confirmed by amplification of a human-rodent monochromosome hybrid panel (Drwinga et al. 1993) (data not shown).

Comparison of the 7q and 12q sequences revealed high levels of sequence similarity (97.9%). Sequence database (BLAST) searches, done with the 12q sequence, revealed a distinct tripartite structure that is unique to both the 7q and 12q telomere–adjacent sequences. The sequence (125 bp) immediately adjacent to the 12q telomere showed homology (86%) to two interstitial telomere-like repeat arrays (fig. 2). These two interstitial telomere-like sequences were within clones derived from 16p11.1 and 2p11 (European Molecular Biology Laboratory (EMBL) Database accession numbers ac002038 and ac002307). Proximal to the 125 bp immediately adjacent to the 12q and 7q telomeres are 512 bp (−125 to −637) with homology to subtelomeric repeat sequences shared by many chromosome ends, including the interstitial telomere-like arrays of 16p and 2p. At least 18 different sequences showed an average sequence similarity of 83% in this region. Homology to these sequences is abruptly terminated (at position −666) by

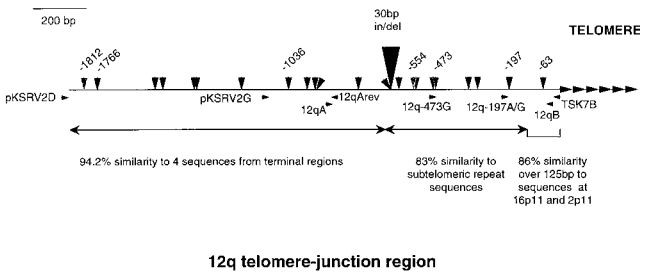

Figure 2 Diagrammatic representation of the 12q telomere–adjacent DNA. The positions of primers used for PCR analysis are denoted by a *horizontal arrow,* and polymorphic sites are denoted by a *vertical arrow.* Polymorphic sites used either in the RFLP analysis or for allele-specific amplification are denoted by their position with respect to the first telomere repeat. Homologies with other subtelomeric sequences are indicated.

Table 1

12q Telomere–Adjacent Haplotype Frequencies, as Defined by RFLP Analysis of the White Population (CEPH)

	Polymorphic Sites[a]							
Haplotype	−1812 *Alu*I	−1036 *Taq*I	−(659–630)[b]	−554 *Kpn*I	−473 *Bsm*I	−63 *Aci*I	No. of Alleles (n = 160)	Frequency
12qA	G	C	+	A	A	C	39	.24
12qA1	G	C	+	G	A	C	28	.18
12qC	G	C	−	A	A	C	1	.01
12qB	A	T	−	A	G	T	60	.38
12qD	A	T	+	A	G	T	1	.01
Deletion	G	NP	NP	NP	NP	T	10	.06
Null	NP	NP	NP	NP	NP	NP	21	.13

[a] NP = sequences are not present.
[b] Bases are present (+) or absent (−), as indicated.

sequences with homology to a less-common subterminal-sequence family (fig. 2). The organization of subterminal-repeat sequences into a two-domain structure has previously been observed over longer tracts of sequences at other chromosome ends in humans and in yeast (Flint et al. 1997). Four sequences from the EMBL Database and GenBank—but not the clones derived from 16p and 2p, mentioned above—showed a 94.2% average sequence similarity to the proximal region of the 12q and 7q telomere–adjacent DNA. In summary, the particular tripartite arrangement of subterminal sequences described above has been detected only adjacent to telomere-repeat arrays at 7q and 12q. This, in addition to the low level of sequence divergence between the 7q and 12q telomere–adjacent sequence, indicates that these telomeres arose from a common ancestral sequence.

Scan for Sequence Polymorphism in the 12q Telomere–Adjacent DNA

Ten random DNA samples from white individuals (CEPH) were scanned for sequence polymorphism, by means of 12q-specific PCR amplification and DNA sequencing. The 12q telomere–adjacent sequence was amplified in two parts, by use of primers 12qA and TSK7B (distal region) and primers 12qArev and pKSRV2.D (proximal region) (fig. 2). A total of 1,870 bp of DNA was scanned, and 20 single-base substitutions and a 30-bp insertion/deletion were detected. The 30-bp in/del has a high CG content (85%) and occurs at the junction between the two subterminal sequence families described above (fig. 2). The polymorphic sites were numbered according to their position from the start of the telomere-repeat array. Interestingly, the individuals sequenced were either multiply homo- or heterozygous at all of the polymorphic sites identified. This suggests that, like the telomere-adjacent DNA of Xp/Yp, the 12q telomere–adjacent DNA exhibited a high level of linkage disequilibrium.

RFLP Analysis

To further investigate this linkage disequilibrium, a population survey of some of the polymorphic sites was undertaken. Six sites that altered the digestion pattern of restriction enzymes were chosen for analysis. These were the −63 (*Aci*I+/−), −473 (*Bsm*I+/−; fig. 1), −554 (*Kpn*I+/−), −(659–630) in/del (assay on the *Bsm*I+/− digest; fig. 1), −1036 (*Taq*I+/−) and −1812 (*Alu*I+/−) sites. Four of the sites (−63, −473, −1036, and −1812) were found to be in complete linkage disequilibrium in the white population, defining two haplotypes that were known as "A" and "B" (with frequencies in the white population of .24 and .38, respectively; table 1). The −554 (*Kpn*I+/−) site defined a common subset of haplotype A alleles (known as "haplotype A1" [frequency .18]) containing a G at this position (*Kpn*I−). The −(659–630) in/del was also in linkage disequilibrium with the −63, −473, −1036, and −1812 sites, but two rare (1/160 alleles; frequency .01) haplotypes (known as "C" and "D") were observed (table 1). Haplotypes C and D differ from haplotypes A and B, respectively, as a result of a switch at the −(659–630) in/del. The mechanism underlying this switch may include a localized patch of "DNA conversion" between alleles or recurrent deletion/duplication of this sequence. The GC-rich duplicated region may be capable of forming hairpin structures (as predicted by use of the Stem-Loop program, version 6.2; Genetics Computer Group). Such structures are prone to instability (Akgun et al. 1997). In summary, this analysis identified two major haplotypes (A and B) with three derivatives—one (A1) that is common and two (C and D) that are rare (table 1).

Definition of 12q Telomere–Adjacent Haplotypes

The RFLP data demonstrated that the base substitutional polymorphisms in this region exhibited strong linkage disequilibrium. To determine whether the intervening 15 polymorphic sites were also in linkage disequilibrium with haplotypes A, B, and A1, individuals who were identified, from the results of the RFLP analysis, as being either completely homo- or hemizygous (see below) were sequenced across the entire region. Amplicons generated from PCR amplification with the 12q-specific primer 12qB and with primer pKSRV2.D were sequenced with primers pKSRV2.D, pKSRV2.G, 12qArev, 12qA, and 12qB (fig. 2). A total of 9 haplotype A, 10 haplotype B, and 9 haplotype A1 alleles were sequenced in this way (see the Material and Methods section). All the polymorphic sites were in linkage disequilibrium, defining the three haplotypes A, A1, and B (see EMBL Database accession numbers AJ132885, AJ132886, and AJ132887, respectively; table 2). Two additional variant positions were identified: one (−646 C→G; 22% of haplotype A1 alleles; estimated population frequency .07), which was present in a subset of haplotype A1 alleles, was contained within the in/del region, and one (−1212 C→G; estimated population frequency .11) was found in a subset (30%) of haplotype B alleles (table 2). Haplotypes A and B showed a sequence divergence of 1.1% in the region between the most proximal polymorphic site (−1812) and the start of the telomere-repeat array. The ancestral sequence for some of the polymorphic positions in the 12q-adjacent sequence was determined by comparison of the 12q haplotypes with those found in related subterminal sequences and in the 7q telomere–adjacent sequence.

Null Alleles

The primers used in the RFLP analysis failed to amplify two DNAs. In addition, the frequencies of the haplotypes were not concordant with those expected of a population in Hardy-Weinberg equilibrium, as there was an excess of homozygous individuals. One of the DNAs could be amplified across the entire region, with use of primers 12qB and pKSRV2.D. The amplification product was ~1.5 kb shorter than was expected (fig. 3*b*). Sequence analysis of this fragment revealed a 1,439-bp deletion that included the annealing sites for primers 12qA and 12qArev, which were used in the amplifications for the RFLP analysis (EMBL Database accession number AJ132888). The population frequency (.06) of this deletion haplotype was determined by PCR analysis of the CEPH parental DNAs, done with the use of 12qB and pKSRV2D primers (fig. 2). No sequence variation was detected among the four deletion alleles that were sequenced. The proximal 297 bp of the deletion allele show homology to the 12q haplotype A sequence, as defined by the −1812 and −1766 polymorphic sites. Distal to the deletion breakpoint, there are 135 bp of homology to the haplotype B sequence, as defined by a single polymorphism (−63 position) (fig. 3*b*). The sequence of the first 37 telomere repeats of one 12q deletion–associated telomere was determined. Surprisingly, the repeat array was dissimilar to those associated with haplotype B. The first eight repeat units were identical to telomeres associated with haplotype A, but they have been duplicated three times in the deletion-associated telomere. The distal 13 repeats that were sequenced were dissimilar to haplotype A telomeres (figs. 3*a* and 4). The telomere-adjacent DNA of the deletion allele contained an additional 23 bp of sequence at the breakpoint (−135 to −1574; fig. 3*c*). The additional sequence contained 15 bp of a 19-bp inverted repeat derived from a sequence located 14 bp distal to the breakpoint; the other 4 bp of the inverted repeat were present immediately distal to the breakpoint. The remaining 8 bp of additional sequence at the breakpoint consisted of a direct repeat (10 bp total) derived from the inverted repeat. Several events must have occurred to generate the complex changes present in the 12q deletion allele, and the lack of variation between the four sequenced copies of the deletion allele suggests either that these events occurred together (at one point in time) or that only a single lineage exists today.

Proximal to the proposed deletion breakpoint, the sequence was more similar to 12q haplotype A than to 12 q haplotype B. However, there were five base differences from the 12q haplotype A sequence located proximal to the breakpoint (fig. 3*c*), and there was another single-base difference (A→C) located 285 bp proximal to the breakpoint. Since three of the five base differences were shared by other copies of the subterminal sequences elsewhere in the genome (fig. 3*c*), it remained possible that the deletion allele was not located on 12q. Therefore, linkage analysis was conducted by use of the CEPH panel of family DNAs. Strong evidence in favor of tight linkage between the 12q telomere–adjacent locus and the distal D12S357 locus (Dib et al. 1996) was found (at recombination fraction [θ] 0, a LOD score of 7.82 was obtained for three families segregating the A, A1, or B alleles at the 12q telomere–adjacent locus and for which segregation data was available at the D12S357 locus [CEPH Genotype Database, version 7]). Segregation analysis also revealed complete concordance between the 12q deletion allele and inheritance at the D12S357 locus (at $\theta = 0$, a LOD score of 3.01 was obtained for two families, when only the segregation data for the 12q deletion allele was included).

The second DNA that did not amplify during the RFLP analysis also failed to amplify with the use of any

Table 2

12q Telomere–Adjacent Haplotypes, as Defined by Sequence Analysis

	Polymorphic Sites																						
	−1812							−1036								−554			−473				−63
Haplotype	*Alu*I	−1766	−1544	−1511	−1391	−1387	−1212	*Taq*I	−963	−934	−933	−769	−(659–630)[a]	−646	−617	*Kpn*I	−550	−482	*Bsm*I	−350	−319	−197	*Aci*I
12qA	G	A	T	C	C	G	C	C	C	T	T	A	+	C	T	A	T	C	A	G	G	A	C
12qA1	G	A	T	C	C	G	C	C	C	T	T	A	+	C/G	T	G	C	C	A	G	G	A	C
12qB	A	G	G	A	A	C	C/G	T	G	A	A	T	−	NP	C	A	C	T	G	C	C	G	T
7q	G	A	C	A	C	C	C	T	C	T	T	T	−	NP	C	A	C	C	G	C	C	G	C
Subterm/prx[b]	G	G	T	C	A/C	C	C	C	C	T	T	T	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
Subterm/dist[c]	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	C	NP	NP	NP	G	C	C	C	C
Ancestral	G		T			C	C		C	T	T	T			C	A	C	C	G	C	C		C

Note.—The numbers represent the position from the first telomere repeat. Five polymorphic sites that create RFLPs are indicated. NP = sequence not present.

[a] The −(659–630) in/del is observed with the *Bsm*I assay for the −473 position; the 30 bases between −659 and −630 are present (plus sign [+]) or absent (minus sign [−]), as indicated.

[b] Subterminal repeat sequences (EMBL Database accession numbers ac002055, x92108, al031259, and x58156) with homology to the proximal region of the 12q telomere–adjacent sequence.

[c] Subterminal repeat sequences (EMBL Database accession numbers ac002307 and ac002038) with homology to the distal region of the 12q telomere–adjacent sequence.

Table 3

Xp/Yp Telomere–Adjacent Haplotypes, as Defined by Sequence Analysis

	Polymorphic Sites																							
Haplotype	−1078	−1027	−989	−848	−842	−826	−652	−554	−544	−540	−427	−415	−(373–363)[a]	−338	−333	−325	−298	−297	−176	−147	−146	−145	−73	−30	−13
Xp/Yp D	C	C	C	T	C	T	G	A	C	C	G	C	−	A	C	C	A	T	G	G	G	C	G	T	T
Xp/Yp A	A	T	G	C	C	C	G	A	G	T	G	C	+	G	A	G	A	T	G	A	G	T	C	T/A	T/A
Xp/Yp B	C	T	G	C	A	C	A	T	C	C	A	T	+	G	A	G	G	C	T	A	A	T	C	A	A
Xp/Yp C	C	T	G	C	A	C	A	A	G	T	G	T	+	G	A	G	G	C	T	A	G	T	C	A	A
Ancestral[b]	C	C	G	C	C	C	G	A	C	T	G	C	+	G/A	C	G	A	T	T	A	G	T	C		

Note.—Polymorphic sites that have been underlined are newly identified on the Xp/Yp haplotype D (previously described as African haplotype A [Baird et al. 1995]).

[a] The 10 bases between −363 and −373 are present (plus sign [+]) or absent (minus sign [−]), as indicated.

[b] Derived from the chimpanzee, gorilla, and orangutan sequences.

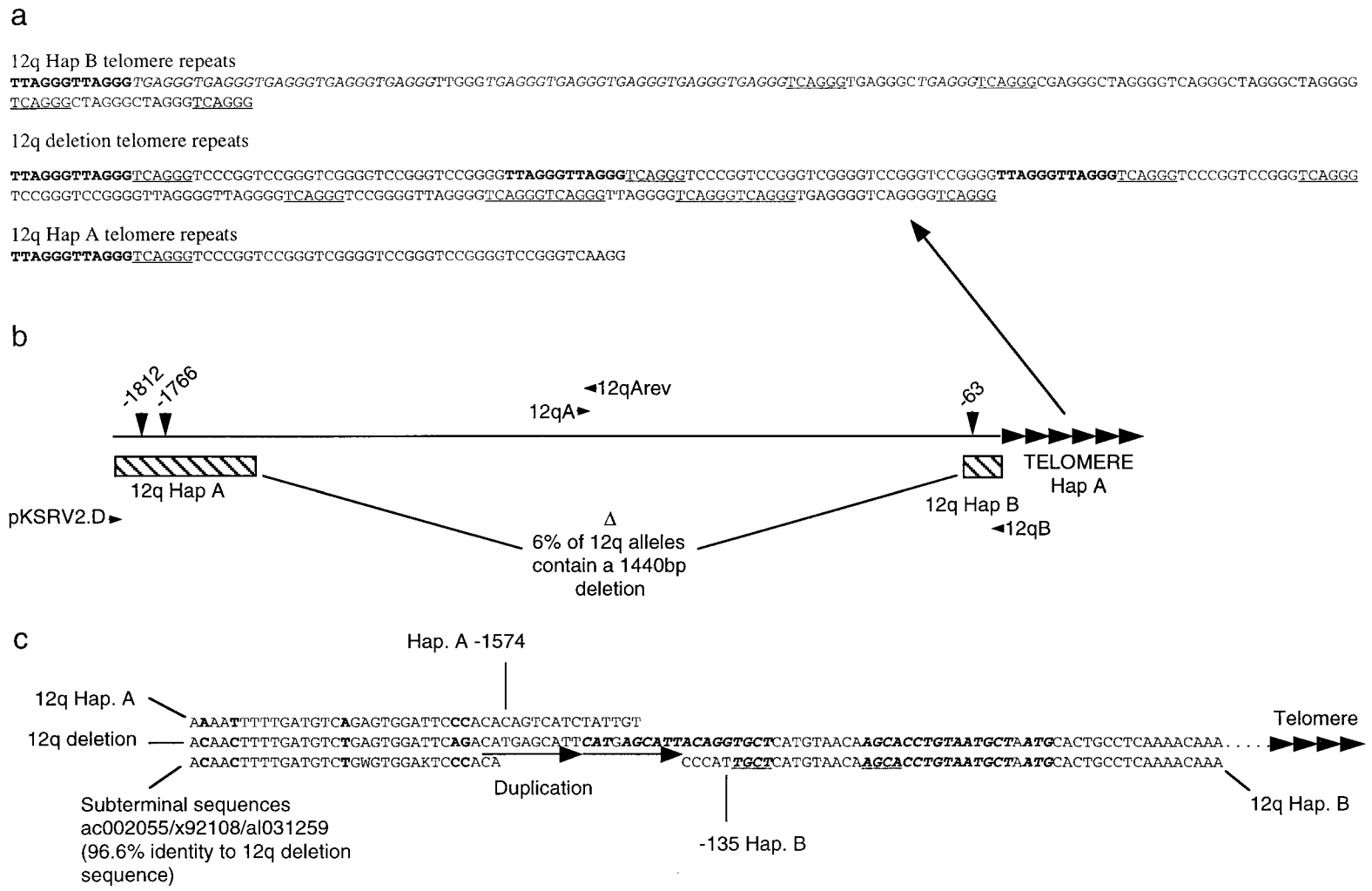

Figure 3 DNA sequence of the 12q deletion allele compared with the sequences of haplotypes A and B. *a*, The sequences of the telomere-repeat arrays from one deletion allele and from haplotype A and B alleles are shown; TTAGGG repeats are indicated by boldface text; TCAGGG repeats, by underlined text; TGAGGG repeats, by italicized text; and other repeats, by plain text. *b*, Diagram representing the deletion allele and telomere-repeat array. *c*, Sequence features of the deletion breakpoint include a 19-bp inverted repeat sequence, which is indicated by bold italicized text; note the single sequence difference (indicated by plain text) between the two copies of the inverted repeat. Also shown are the duplication of the proximal 10 bp of the inverted repeat sequence (*horizontal arrow*), a 4-bp inverted repeat sequence immediately distal to the breakpoint (indicated by underlined text), and sequence differences between the deletion allele and the haplotype A proximal to the breakpoint (indicated by boldface text).

Figure 4 Telomere variant-repeat codes derived from 12q telomeres in the white population. These codes were generated from the −197 polymorphism in individuals who were hetero- or hemizygous for the haplotypes in the 12q telomere–adjacent DNA. For each telomere code, the haplotype in the 12q telomere–adjacent DNA is indicated. Codes were grouped by visual inspection of the patterns, according to coding similarities. Repeat types are represented as follows: dash (-) = TTAGGG; G = TGAGGG; C = TCAGGG; and N = unknown repeat type. The TVR codes from two deletion alleles, Δ, were generated from the deletion-specific primer 12qnull3.

of the 12q-specific primers. This individual presumably contains a 12q telomere with adjacent DNA that is not homologous to the 12q telomere–adjacent alleles described here. To determine the frequency of this "null" allele in the population, segregation analysis was performed in the families of parents scored by CEPH as being homozygous for the 12q haplotypes. The null allele segregated in a Mendelian fashion in pedigrees, and some children who were homozygous for null alleles were identified. Consequently, some parents who were previously designated as being homozygous for one of the 12q haplotypes were identified as 12q null/12q A, A1, or B hemizygotes. The frequency of the null allele was estimated to be .13 in the white population (table 1). When the corrected genotypes were included, the population was shown to be in Hardy-Weinberg equilibrium. Furthermore, the results of linkage analysis performed on seven CEPH families segregating either the 12qA, A1, B, deletion, or null alleles at the 12q telomere–adjacent locus confirmed the close linkage to the D12S357 locus (LOD score = 14.75 at $\theta = 0$).

Telomere- and Variant-Repeat Distribution in the 12q Telomere

Allele-specific PCR primers (12q-197A and 12q-197G) to the −197 A→G polymorphism were designed. When these allele-specific primers are used in conjunction with repeat-specific telomere primers, the interspersion pattern of variant repeats along single telomeres can be determined in individuals considered to be hetero- or hemizygous at the −197 position. The genotype at the −197 position was determined for some individuals (data not shown), but the presence of strong linkage disequilibrium in this region obviated the need to genotype all individuals at the −197 position. For most individuals, the −197 genotype was inferred from the genotypes at the flanking −63 and −473 positions (table 2). The TVR-PCR technique uses repeat primers designed to detect TTAGGG, TGAGGG, and TCAGGG repeat types (fig. 5), and the distribution of repeat types along each array is represented as a four-character code (fig. 4) (Baird et al. 1995).

The repeat distribution at the 12q telomere was found to be hypervariable; 53 different telomere codes were observed among the 53 telomeres mapped. Telomere codes were grouped according to similarities in the repeat-distribution pattern; these grouped alleles were associated with the same haplotype in the DNA adjacent to the telomere (fig. 4). Telomere codes associated with each of the flanking haplotypes possessed specific characteristics.

Codes associated with haplotypes A and A1.—All of the haplotype A– or A1–associated codes ($n = 24$) started with the characteristic TTNNNNNNNNCGNN

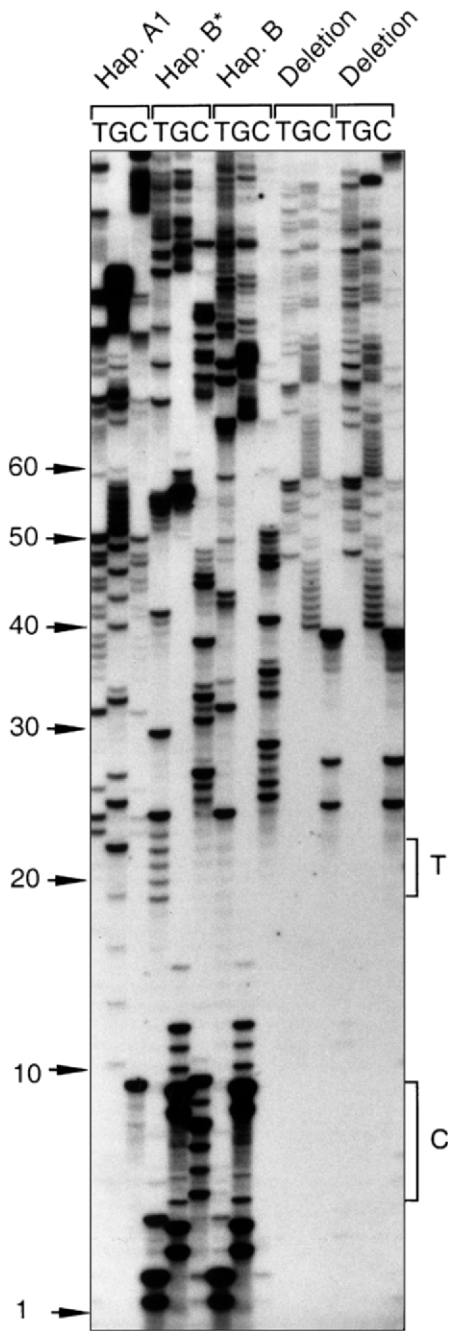

Figure 5 An example of an autoradiograph of the 12q-specific TVR-PCR reactions. Five allele-specific TVR-PCR reactions are shown. Primer 12q-197A was used to amplify haplotype A and A1 alleles specifically; primer 12q-197G, to amplify haplotype B alleles; and 12qdel, to amplify the deletion-associated telomeres. The repeats are numbered from the first repeat in the 12q telomere–repeat array, as shown. The 12q-197G primer is capable of annealing to the 7q telomere–adjacent DNA; consequently, 7q-specific products can be generated with use of this primer. These are observed in the T and C lanes of one HapB telomere (indicated by an asterisk [*]); the bands are bracketed on the right side of the autoradiograph. Repeat types are represented as follows: T = TTAGGG, G = TGAGGG, and C = TCAGGG.

repeat motif (fig. 4), and, further along the array, these telomeres contained a complex interspersion of TGAGGG and TTAGGG repeat types. The haplotype A1–associated codes all (with the exception of 0202) contained the repeat motif TNG reiterated into variable-length arrays. Only one haplotype A code (134602) contained a similar number of the TNG motifs (fig. 4). The similarity at the start of the repeat arrays suggests that the haplotype A– and A1–associated telomeres were derived from a single common ancestral telomere. This finding is consistent with the fact that haplotypes A and A1 differed at only the −550 and −554 *Kpn*I positions in the telomere-adjacent DNA (table 2). Two of the haplotype A–associated telomere codes contained large blocks ($n = 20$ and $n = 51$) of the TCAGGG repeat unit. The majority of the haplotype A– or A1–associated telomeres contained variation beyond the resolution of the electrophoretic systems used. One telomere-repeat code associated with the rare 12q haplotype C was obtained. Since haplotype C differs from haplotypes A and A1 only at the in/del polymorphism (table 1), it was not surprising that its telomere code was very similar to the haplotype A and A1 telomere codes.

Codes associated with haplotype B.—In contrast to the haplotype A– and A1–associated telomeres, the first 60 repeats of the haplotype B–associated telomeres contain less variation between arrays (fig. 4). This region is characterized by three blocks of TCAGGG repeat units, but this repeat type is uncommon in telomeres associated with the other 12q haplotypes. The limited variation observed in this region of the haplotype B–associated telomeres consists of small alterations in blocks of like repeat units. The majority (22/26) of these telomeres contain a duplication of the first 60 repeat units, and, as in the first 60 repeats, the duplicated region shows limited variation. Beyond this duplication is a complex and hypervariable interspersion of TGAGGG and TTAGGG repeat types, but this variation is beyond the resolution of the electrophoretic systems used in the present analysis and has not been characterized in detail (fig. 5). Two telomeres (141601 and 134001; fig. 4) are composed of an unduplicated copy of the first of the first 60 repeat units, and these telomere-repeat-distribution patterns may represent the ancestral telomere (prior to the duplication).

Telomere codes associated with the adjacent 12q deletion allele.—Primer 12qnull3 was designed from the sequence of the deletion breakpoint, to allow allele-specific amplification of the 12q deletion–associated telomere. Two similar codes were derived from 12q deletion–associated telomeres (figs. 4 and 5), and these telomeres started with the TTNNNNNN motif observed at the beginning of all haplotype A– and A1–associated telomeres. This motif has been duplicated twice. Distal to this duplicated region is a complex interspersion of TGAGGG and TTAGGG repeats.

Extension of Sequence Analysis of Xp/Yp Telomere–Adjacent Haplotype

The results of a previous analysis of sequence polymorphism and linkage disequilibrium at the Xp/Yp telomere revealed two highly diverged haplotypes (A and B) in white individuals with a third common haplotype, C, that was very similar to haplotype B. Sequence analysis of these haplotypes in Africans revealed additional heterogeneity in haplotype A (Baird et al. 1995). To calculate sequence divergence between the Xp/Yp telomere–adjacent haplotypes accurately, it was necessary to investigate the additional heterogeneity identified in the African population. Furthermore, TVR-PCR analysis of African haplotype A–associated telomeres showed that, in contrast to haplotype A–associated telomeres derived from white individuals, the African haplotype A telomere codes were more diverse and could be divided into two major groups (fig. 6). One group, which began with a block of NNG_{6-7} repeat types, was very similar to that observed in haplotype A–associated codes in white individuals (Baird et al. 1995). The other group of telomere codes was quite distinct (fig. 6). To determine whether this distinct group of African telomere codes was associated with the additional variation in the telomere-adjacent DNA, sequence analysis of the adjacent DNA was done. Telomere-adjacent DNAs, obtained from two individuals (AF28 and 10202), that contained telomere codes from the distinct group (fig. 6) were sequenced. The sequence of the telomere-adjacent DNA of these two telomeres was identical and was different from the previously defined Xp/Yp haplotype A (table 3). It was designated as "haplotype D." It seems likely that the African telomeres shown in figure 6 can be subdivided into two groups—one that is similar to the flanking haplotype A–associated telomere codes in white individuals and a second group that is associated with the newly defined flanking haplotype D. One of the sequenced haplotype D alleles, derived from a white individual (10202; fig. 6), contained a telomere code similar to the distinct group of African codes. Haplotype D appears to be uncommon in white individuals, but it is relatively common in Africans and represents additional variation in the Xp/Yp telomere–adjacent DNA. Nevertheless, the 25 polymorphic sites in the 1,078 bp of DNA immediately adjacent to the Xp/Yp telomere–repeat array are in strong linkage disequilibrium, since only four distinct haplotypes have been observed (table 3). These haplotypes exhibit sequence divergence of $\leq 2.0\%$ (between haplotypes D and B). Intermediate haplotypes, which might have been expected if the sequence divergence had arisen by sequential mutations,

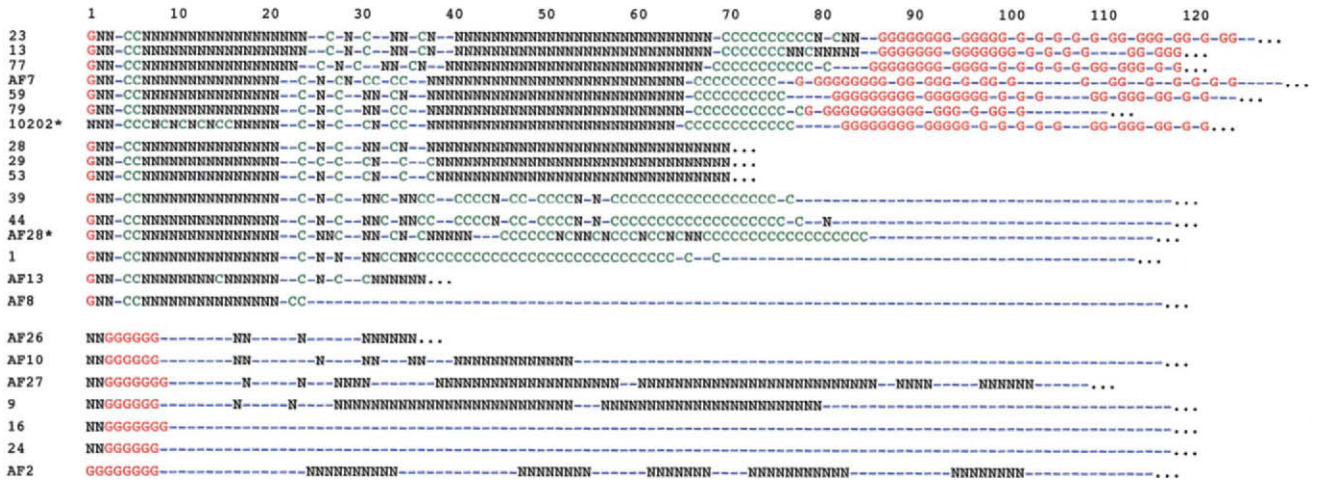

Figure 6 Xp/Yp-telomere codes derived from 22 Africans and from one white individual. Two types of TVR codes were observed. One, which starts with the characteristic repeat motif GNNTCCNNNN…, is common in the African population and is associated with the telomere-adjacent haplotype D in the two individuals (10202 and AF28, who are denoted by an asterisk [*]) analyzed. Individual 10202 is white. The other African TVR codes that are shown start with the NNGGGGGG repeat motif that is characteristic of the Xp/Yp haplotype A–associated TVR codes in white individuals. The different repeat types are represented by the same characters that are used in figure 4.

have not been identified for the majority of the polymorphic positions. An exception to this is seen in haplotype C, which differs from haplotype B between the −554 and −427 polymorphisms and at the −146 polymorphism (table 3).

Discussion

Telomere-Variant-Repeat Coding

In this article we have described the patterns of sequence polymorphism within both the telomere of 12q in the white population and the subset of Xp/Yp telomeres in the African population. The distribution of telomere and variant repeats at 12q is hypervariable, as described in other telomeres (Baird et al. 1995; Coleman et al. 1999). Most of the variation occurs as the result of complex interspersion patterns of TTAGGG and TGAGGG repeats. Despite the variation between alleles, it is apparent that subsets of 12q alleles share similar TVR codes. Each subset of alleles must have evolved as a haploid lineage from a recent ancestral TVR code. Small localized expansions and contractions in blocks of like repeat units or in the copy number of a higher-order repeat motif can account for differences between closely related alleles. In general, both TCAGGG and null repeats are confined to the first 100 repeats of the 12q telomere, whereas the complex interspersion pattern involving TGAGGG repeats can extend ⩾200 repeats into the telomere-repeat array. There may have been insufficient evolutionary time, since the relatively recent appearance of human telomeres at their current locations, for variant repeats to spread uniformly through the proximal ends of telomeres. However, the different distributions of the TGAGGG and TCAGGG repeat types have been observed at the Xp/Yp and 12q telomeres and may reflect functional constraints acting to reduce the spread of some variant repeats more than others. Such constraints may take the form of either altered chromatin structure or the different affinities of telomere binding proteins, such as TRF1, for particular repeat types (Chong et al. 1995; Bianchi et al. 1997). Alternatively, the presence of t-loops, formed when the 3′ terminus of the telomere is tucked back into the proximal regions of the telomere, perhaps near the telomere-adjacent sequences, might serve to limit the spread of variant repeats (Griffith et al. 1999). Functional constraints such as these may also explain the apparent conservation of the TTAGGG telomere repeat at the proximal ends of telomeres, compared with the 12q-related interstitial telomere-like sequences at 16p11.1 and 2p11, which are composed mostly of such highly degraded repeats as TTAGGAGTCAACGTTTAGAG. Furthermore, if true, this suggests that these variant repeats may be capable of supporting some of the functions performed by TTAGGG telomere repeats.

None of the TVR codes analyzed were identical, but the 12q haplotype B–associated alleles were more homogeneous than the other telomeres examined to date. The majority of telomeres associated with the 12q haplotype B contained a duplication of the first 60 repeat units, and it may be that this particular arrangement of repeats exerts a stabilizing effect.

High Level of Sequence Polymorphism in the Telomere-Adjacent DNA

The unusual genetic properties of the telomere-adjacent DNA at the Xp/Yp telomere (Baird and Royle 1997; Baird et al. 1995) are not unique to this telomere, since the telomere-adjacent DNA of 12q also exhibits high levels of sequence polymorphism and linkage disequilibrium. A total of 22 SNPs and one 30-bp insertion/deletion polymorphism were identified in 1,870 bp of DNA adjacent to the 12q telomere. This represents a frequency of one SNP/85 bp, which is much higher than the genome average of one SNP/0.3–1 kb (Cooper et al. 1985; Wang et al. 1998; Cargill et al. 1999), although it is lower than the frequency of one SNP/45 bp in the Xp/Yp telomere–adjacent DNA. The results of analysis of the sequences orthologous to the human Xp/Yp telomere–adjacent DNA in great apes have shown that these sequences exhibit higher levels of sequence divergence ($\times 2.6$) than do other genomic loci (Baird and Royle 1997). In addition, the Xp/Yp telomere–adjacent sequences in orangutans also displayed much sequence variation, and two highly diverged haplotypes (2% divergence between the haplotypes) were identified. These observations indicate that sequences adjacent to telomeres in humans and orangutans display elevated levels of mutation and that it is possible that an unusual chromatin conformation in these regions reduces the accessibility of repair proteins to the telomere-adjacent DNA. This does not, however, explain the presence of linkage disequilbrium across the polymorphisms in the telomere-adjacent DNA at Xp/Yp and 12q.

Linkage Disequilibrium at the Xp/Yp and 12q Telomeres

The strong linkage disequilibrium in the telomere-adjacent DNA of Xp/Yp and 12q results in a limited number of highly diverged haplotypes extending into the telomere-repeat arrays at both the Xp/Yp and 12q telomeres. Recently, the results of a comprehensive analysis involving simulations of a human population showed that linkage disequilibrium between two selectively neutral loci that are in the presence of genome-average levels of recombination is likely to be at a maximum when the two loci are separated by a very small physical distance (~300 bp). The linkage disequilibrium rapidly declined to below "useful" levels if the physical distance was >3 kb (Kruglyak 1999), and it was shown that the level of linkage disequilibrium between pairs of very closely linked loci (separated by 300 bp) was very variable. Obviously, the level of linkage disequilibrium between pairs of loci depends on the population history and can be influenced either by factors such as a severe reduction in the effective population size (Ne) or a reduction in Ne for many generations or by allelic frequencies prior to a reduction in Ne (bottleneck). Therefore, the observation of strong linkage disequilibrium between pairs of very closely linked loci (<3 kb) does not necessarily indicate that the recombination rate is below that of the genome average but, rather, that it may be a reflection of population history. The linkage disequilibrium observed in telomere-adjacent sequences (Xp/Yp and 12q) encompasses multiple SNP loci (one SNP/45 bp and one SNP/85 bp, respectively) within a short physical distance. Detailed haplotype analysis, done over a 3-kb sequence surrounding the β-globin gene, revealed a dense clustering of SNP loci in a region regarded as a recombination hotspot with no evidence of linkage disequilibrium. However, linkage disequilibrium was notable between SNP loci located downstream of this recombination hotspot (Harding et al. 1997). The history of early human populations is unknown or, at least, uncertain; therefore, overall, the presence of strong linkage disequilibrium across multiple SNPs at two telomere-adjacent sequences and extending into the telomeres indicates that recombination is suppressed at these loci. The distal limit of the suppression of recombination within the telomeres is unknown. More sequence analysis is required from the 12q telomere–adjacent DNA, to establish the proximal limit of the suppression of recombination at this locus; however, it extends for only ~1 kb into the telomere-adjacent DNA at Xp/Yp (Baird and Royle 1997). Linkage-map expansion and chiasmata distribution in the terminal regions of human chromosomes indicate an enhancement of recombination in these regions (Hulten 1974; National Institutes of Health/CEPH 1992; Mohrenweiser et al. 1998). In addition, evidence from yeast demonstrates that unequal recombination between nonhomologous telomeres can provide a mechanism for telomere maintenance in the absence of telomerase activity (Lundblad and Blackburn 1993). It has been postulated that a similar mechanism may be important in telomere maintenance in human cell lines and cancers containing long telomeres in the apparent absence of telomerase activity (Reddel et al. 1997). The data presented here suggest that human telomeres may have a localized suppressive effect on recombination—at least in the human germline, but this does not preclude the possibility of there being unequal exchanges between telomeres in neoplastic cells, where strong selection would exist for telomere maintenance. In contrast, suppression of recombination between homologous telomeres in the germline might avoid the unnecessary introduction of length heterogeneity between telomeres. A reduction in recombination rate has been observed in yeast as a result of the telomere-position effect, and it is thought to be dependent upon specific

telomere binding proteins that form the telosome structure of yeast telomeres (Stavenhagen and Zakian 1998). Similarly, centromeres are known to exert a suppressive effect on meiotic recombination (Mahtani and Willard 1998). The suppression of recombination at human telomeres may be a consequence either of the chromatin structure created by such telomere binding proteins as TRF1 (Chong et al. 1995; Bianchi et al. 1997; Tommerup et al. 1994) or of the presence of t-loops (Griffith et al. 1999). If short telomeres have an altered chromatin structure, the suppressive chromatin effect on recombination may be lost and the short telomere may become available for elongation by means of recombination-based mechanisms (Bryan et al. 1995; Bryan et al. 1997).

Formation of Highly Diverged Haplotypes in Telomere-Adjacent DNA

The presence of high levels of recombination in the proterminal regions of human chromosome effectively isolates human telomeres from the influences of selection operating on the rest of the genome. Therefore, provided that the telomere-adjacent sequences in the genome remain recombinationally suppressed, diverged haplotypes that arise in these regions may persist for a very long time. Suppression of recombination may in part explain the maintenance of a few diverged haplotypes adjacent to the Xp/Yp and 12q telomeres, and an elevated mutation rate could explain the high levels of sequence divergence both within and between species. However, these alone do not explain the apparent lack of intermediate haplotypes that must have existed as a consequence of the sequential mutations required to create the diverged haplotypes. If the intermediate haplotypes have been lost through rounds of mutation and fixation resulting from random genetic drift at neutral loci (Takahata 1996), then the question arises as to why two highly diverged haplotypes persist at the two telomeres. Although there must be strong selection to maintain a telomere, it is difficult to imagine how this could result in the generation and maintenance of diverged haplotypes in the telomere-adjacent DNA of two telomeres. It therefore seems unlikely that the diverged haplotypes arose and that they have been maintained by such selection-based mechanisms as selective sweeps, frequency-dependent selection, and overdominance (Bodmer 1972; Nei 1988; Slade and McCallum 1992; Amos and Harwood 1998).

One explanation is that the diverged haplotypes arose in a single lineage that has undergone a population bottleneck (Harpending et al. 1998). Prior to the bottleneck, the population would have contained an extraordinarily large number of haplotypes in the telomere-adjacent DNA of both Xp/Yp and 12q. To maintain that level of haplotype diversity, the population would have to have been very large and/or the base substitution rate would have to have been very high. The data from studies of mitochondrion variation tend to suggest that there has been a bottleneck during recent human evolution, but the results of studies of nuclear loci do not always support this hypothesis (Clark et al. 1998; Harpending et al. 1998; Zietkiewicz et al. 1998). For example, analysis of a 3-kb region encompassing the β-globin gene did not reveal evidence for a bottlenecked founding population, but it did show evidence for coalescence of the allelic lineages to a common ancestor ~800,000 years ago (Harding et al. 1997).

Two alternative explanations for the presence of divergent haplotypes adjacent to two telomeres can be envisaged. First, the divergent haplotypes arose independently at separate subterminal loci within an archaic hominoid genome. The high level of exchange between subterminal repeat sequences then resulted in the relocation of one of the subterminal sequences with a telomere to the end of the same chromosome, thus creating two highly diverged haplotypes at one locus. We think, however, that this explanation is unlikely, since there is no evidence that "donor" loci exist in the modern genome. The results of linkage analysis indicate that the only copies of the sequences that can be amplified by the 12qA, 12qB, and 12qArev primers are linked to the end of 12q. Also, although a related copy of the 12q telomere–adjacent sequence is present on some copies of chromosomes 7q, the sequence in this location does not show more similarity to one 12q telomere–adjacent haplotype than to the other. In addition, there is no evidence that a second locus with homology to the Xp/Yp telomere–adjacent sequence is present in the human or in other great-ape genomes. It would therefore be necessary to assume that the "donor" loci for the ends of both chromosomes were present in an ancestral genome but have been lost from the modern human genome. Another explanation is that the diverged haplotypes arose, in separate archaic hominoid lineages, from a common ancestral sequence. These lineages would have to have been isolated for sufficient time to allow divergent haplotypes to arise as a result of sequential mutations and of fixation in each lineage for one predominant haplotype. The degree of sequence divergence between the haplotypes would be dependent on the mutation rates of the loci examined. The high mutation rate in the telomere-adjacent sequences would have resulted in rapid divergence of these sequences in the different lineages. A subsequent hybridization of two hominoid lineages would bring the highly diverged haplotypes together. The continued existence of the diverged haplotypes after the hybridization event would depend on factors such as recombination, drift, and founder effect, and it could vary between loci. This model implies that *Homo sapiens*

may have evolved from a hybridization event between two hominoid lineages.

The most detailed information on the evolution of telomere-adjacent sequences is available at the Xp/Yp telomere. Haplotype analysis of the Xp/Yp telomere-adjacent sequence has been conducted in a number of different human populations; the ancestral haplotype has been determined from the orthologous sequences in the chimpanzee, gorilla, and orangutan; and the rate of sequence evolution can be estimated. Haplotypes B and D, present at the human Xp/Yp telomere, exhibit a sequence divergence of 1.9% (between positions −75 and −1080), and the sequence divergence between humans (haplotype B) and chimpanzees, over the same orthologous sequence, is 4.6% (Baird and Royle 1997). The divergence of the *Homo* and *Pan* lineages has been dated to ~4.5 million years ago (Miyamoto et al. 1988; Horai et al. 1995; Takahata and Satta 1997). By use of this information, it is possible to estimate that, for such diverged haplotypes to arise, the required separation time of two lineages is 1.9 million years (4.5/4.6∗1.9).

Identification of when the proposed hybridization event might have occurred is difficult. If the hybridization brought together only the two most highly diverged haplotypes at Xp/Yp (D and B), sufficient time must have elapsed, since the hybridization, for haplotypes A and C to be formed. Alternatively, the hybridization could have occurred more recently, bringing together haplotypes A, B, and C from one lineage and haplotype D from the other. It is likely, however, that some haplotype-specific polymorphisms have arisen since the hybridization event or, indeed, since a population bottleneck. For example, at 12q, polymorphisms specific to the 12qA1 (−646) and 12qB (−1212) haplotypes (table 2) have been detected during the sequence analysis. Since the timing of the proposed hybridization event is unknown, it is not possible to suggest which hominoid lineages may have been involved; however, the common ancestor to the lineages must have existed >2 million years ago, perhaps coinciding with one of the *Australopithecine* species. Additional analysis of the 12q telomere and its adjacent sequence in other human populations may distinguish between the explanations outlined above, but it is not unreasonable to suggest that hybridization between lineages separated by 1.9 million years could occur, because the orangutan subspecies *Pongo pygmaeus abelii* and *Pongo pygmaeus pygmaeus* are capable of producing fertile offspring, despite the existence of molecular data that suggests that the two subspecies diverged 1.5–1.7 million years ago (Zhi et al. 1996). Of the two diverged haplotypes in the orangutan Xp/Yp telomere–adjacent sequence (discussed above), one haplotype (orang-lower) was detected in both the orangutan subspecies, but the second haplotype (orang-upper) was detected only in the *Pongo pygmaeus abelii* subspecies (2/18 alleles) (Baird and Royle 1997; Baird, unpublished data). Furthermore, the observation of homoplasy in skeletons of the *Australopithecine* species *A. africanus* and *A. afarensis* suggests that human evolution was more complex than is currently understood (McHenry and Berger 1998*a;* McHenry and Berger 1998*b*), and, recently, a skeleton with both human and Neanderthal characteristics was identified (Duarte et al. 1999).

One prediction of the scheme of human evolution presented here is that the unusual genetic characteristics described at two telomeric loci may occur elsewhere in the human genome. However, such regions are only likely to be detected in areas of the genome that have reduced levels of recombination and selection and that are isolated from the rest of the genome by recombination hotspots. One possible example is at the hypervariable minisatellite MS32 (Jeffreys et al. 1998*a;* 1998*b*), where polymorphisms in the DNA upstream of the recombination hotspot are in strong linkage disequilibrium.

Acknowledgments

We thank Alec Jeffreys, Mark Jobling, Celia May, Yuri Dubrova, Philippe Bois, and Jerome Buard, for discussions. We also thank Stephen J. O'Brien, for the DNA samples from the two orangutan subspecies. This work was supported by the Medical Research Council of Great Britain.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

EMBL Database, http://helix.nih.gov/science/embl.html (for haplotype A [accession number AJ132885], haplotype A1 [accession number AJ132886], haplotype B [accession number AJ132887], and the sequence of the 12q deletion allele [accession number AJ132888])
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/index.html
CEPH Genotype Database, http://www.cephb.fr/

References

Akgun E, Zahn J, Baumes S, Brown G, Liang F, Romanienko PJ, Lewis S, et al (1997) Palindrome resolution and recombination in the mammalian germ line. Mol Cell Biol 17:5559–5570

Allshire RC, Dempster M, Hastie ND (1989) Human telomeres contain at least three types of G-rich repeat distributed nonrandomly. Nucleic Acids Res 17:4611–4627

Amos W, Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. Philos Trans R Soc Lond B Biol Sci 353:177–186

Baird DM, Jeffreys AJ, Royle NJ (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable

variant repeat distribution patterns in the human Xp/Yp telomere. EMBO J 14:5433–5443

Baird DM, Royle NJ (1997) Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. Hum Mol Genet 6:2291–2299

Bianchi A, Smith S, Chong L, Elias P, de Lange T (1997) TRF1 is a dimer and bends telomeric DNA. EMBO J 16:1785–1794

Blackburn EH (1991) Structure and function of telomeres. Nature 350:569–573

Bodmer WF (1972) Evolutionary significance of the HLA system. Nature 237:139–145

Bryan TM, Englezou A, Dalla-Pozza L, Dunham MA, Reddel RR (1997) Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nat Med 3:1271–1274

Bryan TM, Englezou A, Gupta J, Bacchetti S, Reddel RR (1995) Telomere elongation in immortal human cells without detectable telomerase activity. EMBO J 14:4240–4248

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, et al (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238

Chong L, Van Steensel B, Broccoli D, Erdjument-Bromage H, Hanish J, Tempst P, De Lange T (1995) A human telomeric protein. Science 270:1663–1667

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, et al (1998) Haplotype structure and population genetic inferences from nucleotide- sequence variation in human lipoprotein lipase. Am J Hum Genet 63: 595–612

Coleman J, Baird DM, Royle NJ (1999) The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q. Hum Mol Genet 8:1637–1646

Coleman J, Royle NJ (1996) Telomere variant repeat mapping used to reveal variation at human autosomal telomeres. Am J Hum Genet 59 SupplA:176

Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J (1985) An estimate of unique DNA-sequence heterozygosity in the human genome. Hum Genet 69:201–205

de Lange T, Shiue L, Myers RM, Cox DR, Naylor SL, Killery AM, Varmus HE (1990) Structure and variability of human-chromosome ends. Mol Cell Biol 10:518–527

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Drwinga HL, Toji LH, Kim CH, Greene AE, Mulivor RA (1993) NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. Genomics 16:311–314

Duarte C, Mauricio J, Pettitt PB, Souto P, Trinkaus E, van der Plicht H, Zilhao J (1999) The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. Proc Natl Acad Sci USA 96:7604–7609

Flint J, Bates GP, Clark K, Dorman A, Willingham D, Roe BA, Micklem G, et al (1997) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. Hum Mol Genet 6:1305–1313

Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, deLange T (1999) Mammalian telomeres end in a large duplex loop. Cell 97:503–514

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, et al (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Harley CB, Futcher AB, Greider CW (1990) Telomeres shorten during ageing of human fibroblasts. Nature 345:458–460

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Hastie ND, Dempster M, Dunlop MG, Thompson AM, Green DK, Allshire RC (1990) Telomere reduction in human colorectal carcinoma and with ageing. Nature 346:866–868

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Hulten M (1974) Chiasma distribution at diakinesis in the normal human male. Hereditas 76:55–78

Ijdo JW, Lindsay EA, Wells RA, Baldini A (1992) Multiple variants in subtelomeric regions of normal karyotypes. Genomics 14:1019–1025

Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204–209

Jeffreys AJ, Murray J, Neumann R (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Molecular Cell 2: 267–273

Jeffreys AJ, Neil DL, Neumann R (1998) Repeat instability at human minisatellites arising from meiotic recombination. EMBO J 17:4147–4157

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144

Lundblad V, Blackburn EH (1993) An alternative pathway for yeast telomere maintenance rescues *est1*- senescence. Cell 73:347–360

Mahtani MM, Willard HF (1998) Physical and genetic mapping of the human X chromosome centromere: repression of recombination. Genome Res 8:100–110

McHenry HM, Berger LR (1998*a*) Body proportions in *Australopithecus afarensis* and *A. africanus* and the origin of the genus *Homo*. J Hum Evol 35:1–22

——— (1998*b*) Limb lengths in *Australopithecus* and the origin of the genus *Homo*. S Afr J Sci 94:447–450

Miyamoto MM, Koop BF, Slightom JL, Goodman M, Tennant MR (1988) Molecular systematics of higher primates: genealogical relations and classification. Proc Natl Acad Sci USA 85:7627–7631

Mohrenweiser HW, Tsujimoto S, Gordon L, Olsen AS (1998) Regions of sex-specific hypo- and hyper-recombination identified through integration of 180 genetic markers into the metric physical map of human chromosome 19. Genomics 47:153–162

National Institutes of Health/CEPH Collaborative Mapping

Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:148–162

Nei M (1988) Relative roles of mutation and selection in the maintenance of genetic variability. Philos Trans R Soc Lond B Biol Sci 319:615–629

Reddel RR, Bryan TM, Murnane JP (1997) Immortalized cells with no detectable telomerase activity: a review. Biochemistry (Mosc) 62:1254–1262

Royle NJ (1995) The proterminal regions and telomeres of human chromosomes. In: Hall JC, Dunlap JC (eds) Advances in genetics, vol. 32. Academic Press, San Diego, pp 273–315

Royle NJ, Baird DM, Jeffreys AJ (1994) A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. Nat Genet 6:52–56

Royle NJ, Hill MC, Jeffreys AJ (1992) Isolation of telomere junction fragments by anchored polymerase chain-reaction. Proc R Soc Lond B Biol Sci 247:57–61

Slade RW, McCallum HI (1992) Overdominant vs. frequency-dependent selection at MHC loci. Genetics 132:861–864

Stavenhagen JB, Zakian VA (1998) Yeast telomeres exert a position effect on recombination between internal tracts of yeast telomeric DNA. Genes Dev 12:3044–3058

Takahata N (1996) Neutral theory of molecular evolution. Curr Opin Genet Dev 6:767–772

Takahata N, Satta Y (1997) Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc Natl Acad Sci USA 94:4811–4815

Tommerup H, Dousmanis A, de Lange T (1994) Unusual chromatin in human telomeres. Mol Cell Biol 14:5777–5785

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, et al (1998) Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. Science 280:1077–1082

Wilkie AO, Higgs DR, Rack KA, Buckle VJ, Spurr NK, Fischel-Ghodsian N, Ceccherini I, et al (1991) Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. Cell 64:595–606

Zhi L, Karesh WB, Janczewski DN, Fraziertaylor H, Sajuthi D, Gombek F, Andau M, et al (1996) Genomic differentiation among natural populations of orang-utan (*Pongo pygmaeus*). Curr Biol 6:1326–1336

Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, et al (1998) Genetic structure of the ancestral population of modern humans. J Mol Evol 47:146–155